A Comprehensive Epithelial Tubo-Ovarian Cancer Risk Prediction Model Incorporating Genetic and Epidemiological Risk Factors

Andrew Lee*[1], Xin Yang*[1], Jonathan Tyrer[2], Aleksandra Gentry-Maharaj[3], Andy Ryan[3], Nasim Mavaddat[1], Alex P. Cunningham[1], Tim Carver[1], Stephanie Archer[4], Goska Leslie[1], Jatinderpal Kalsi[5,6], Faiza Gaba[7], Ranjit Manchanda[7,8], Simon A. Gayther[9], Susan J. Ramus[10,11], Fiona M. Walter[4], Marc Tischkowitz[12], Ian Jacobs[13], Usha Menon[3], Douglas F. Easton[1,2], Paul P.D. Pharoah[1,2], Antonis C. Antoniou[†1]

[1] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, The University of Cambridge, Strangeways Research Laboratory, Cambridge CB1 8RN, UK.

[2] Centre for Cancer Genetic Epidemiology, Department of Oncology, The University of Cambridge, Cambridge CB1 8RN, UK.

[3] MRC Clinical Trials Unit at UCL, Institute of Clinical Trials & Methodology, University College London, London, WC1V 6LJ UK.

[4] The Primary Care Unit, Department of Public Health & Primary Care, The University of Cambridge, Cambridge, CB1 8RN, UK.

[5] Department of Women's Cancer, Institute for Women's Health, University College London, London, WC1E 6HU, UK

[6] UCL Research Department of Epidemiology & Public Heath, University College London, London, WC1E 7HB

[7] Wolfson Institute for Preventive Medicine, CRUK Barts Cancer Centre, Queen Mary University of London, London, EC1M 6BQ, UK.

[8] Department of Gynaecological Oncology, Barts Health NHS Trust, Royal London Hospital, London E1 1BB, UK.

[9] Center for Bioinformatics and Functional Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[10] School of Women's and Children's Health, Faculty of Medicine, University of New South Wales, Sydney 2052, Australia.

[11] Adult Cancer Program, Lowy Cancer Research Centre, University of NSW, Sydney 2052, Australia.

[12] Department of Medical Genetics and National Institute for Health Research, Cambridge Biomedical Research Centre, The University of Cambridge, Cambridge, UK.

[13] University of New South Wales, Sydney 2052, Australia.

[*]joint first authors

[†]**Corresponding author**: aca20@medschl.cam.ac.uk, +44 1223 748630

Abstract

Background: Epithelial tubo-ovarian cancer (EOC) has high mortality partly due to late diagnosis. Prevention is available but may be associated with adverse effects. Several genetic and epidemiological risk factors (RFs) for EOC have been identified. A multifactorial risk model can help identify females at higher risk who could benefit from targeted screening and prevention.

Methods: We developed an EOC risk model incorporating the effects of family history (FH), pathogenic variants (PVs) in *BRCA1, BRCA2*, *RAD51C*, *RAD51D* and *BRIP1*, a polygenic risk score (PRS) and the effects of RFs. The model was validated in a nested case-control sample of 1961 females from UKCTOCS (374 incident cases).

Results: Estimated lifetime risks in the general population vary from 0.5% to 4.6% for the 1st to 99th percentiles of the EOC risk-distribution. The corresponding range for females with an affected first-degree relative is 1.9% to 10.3%. RFs provided the widest distribution followed by the PRS. In the external validation, absolute and relative 5-year risks were well-calibrated in quintiles of predicted risk.

Conclusion: This multifactorial risk model can facilitate stratification, in particular among females with FH of cancer and/or moderate- and high-risk PVs. The model is available via the CanRisk Tool (www.canrisk.org).

Keywords: ovarian cancer; risk prediction; rare-variants; polygenic risk score; model

Introduction

Epithelial tubo-ovarian cancer (EOC), the seventh most common cancer in females in the world, is often diagnosed at a late stage and is associated with high mortality. There were 7,443 new cases of EOC and 4,116 deaths from EOC annually in the UK in 2015-2017[1]. Early detection could lead to an early-stage diagnosis, enabling curative treatment and reductions in mortality. Annual multimodal screening using a longitudinal serum CA125 algorithm in women from the general population resulted in significantly more women diagnosed with early-stage disease but without a significant reduction in mortality[2]. Four-monthly screening using the same multimodal approach also resulted in a stage shift in women at high risk (>10% lifetime risk of EOC)[3]. Currently, risk-reducing bilateral salpingo-oophorectomy (RRSO), upon completion of their families, remains the most effective prevention option[4] and it has been recently suggested that RRSO would be cost-effective in postmenopausal females at >4% lifetime EOC risk[5,6]. In addition to surgical risk, bilateral oophorectomy may be associated with increased cardiovascular mortality[7] and a potential increased risk of other morbidities such as parkinsonism, dementia, cardiovascular disease and osteoporosis[8,9], particularly in women who do not to take HRT[10]. Therefore, it is important to target such prevention approaches to those at increased risk, who are most likely to benefit.

Over the last decade, there have been significant advances in our understanding of susceptibility to EOC. After age, family history (FH) is the most important risk factor for the disease. Approximately 35% of the observed familial relative risk (FRR) can be explained by rare pathogenic variants (PVs) in 9 genes: *BRCA1*, *BRCA2, RAD51C*, *RAD51D*, *BRIP1*, *PALB2*, *MLH1*, *MSH2* and *MSH6*[11-16]. Common variants, each of small effect, identified through genome-wide association studies (GWAS)[17,18], explain a further 4%. Several epidemiological

risk factors (RFs) are also known to be associated with EOC risk. Use of menopausal hormone therapy (MHT), higher body mass index (BMI), history of endometriosis, late age at menopause and smoking are associated with an increased risk of EOC[19-24], whereas the use of oral contraception, tubal ligation, higher parity and history of breastfeeding are associated with a reduced risk of EOC[22,25,26]. Despite these advances, females at high risk of developing EOC are currently identified mainly through FH of the disease or on the basis of having PVs in *BRCA1* and *BRCA2.* However, more specific risk prediction could be achieved by combining data on all known epidemiological and genetic risk factors. The published EOC prediction models consider either RFs[22,23,27] or common variants[22,28]. No published EOC risk prediction model takes into account all the known EOC susceptibility genetic variants (rare and common), residual cancer FH and the lifestyle and hormonal risk factors.

Using complex segregation analysis, we previously developed an EOC risk prediction algorithm that considered explicit FH of ovarian (and breast cancer), the effects of PVs in *BRCA1* and *BRCA2* and 17 common genetic variants identified through GWAS[11]. The algorithm also modelled the residual, unexplained familial aggregation in terms of a polygenic model that captured the effects of other unobserved genetic effects. The model did not include the effects of other intermediate-risk PVs in genes such as *RAD51C*, *RAD51D* and *BRIP1*[12,13,15,29], which are now included on routine gene panel tests, or the effects of known RFs. Furthermore, no validation in independent datasets was performed.

Here we have extended this model to incorporate the explicit effects of PVs in *RAD51C*, *RAD51D* and *BRIP1*, an up-to-date polygenic risk score (PRS) based on 36 variants and the known EOC RFs (Table s1). We evaluated the performance of this model in the UK

Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)[2], where women from the general population were followed up prospectively.

Methods

EOC risk prediction model development

To develop the model, we used a synthetic approach, previously described in[30] to extend our previous EOC model[11]. In the extended model, the EOC incidence at age t for individual i is given by:

$$\lambda^{(i)}(t) = \lambda_0(t)\exp\left(\sum_{\mu=1}^{6}\left(\beta_{MG\mu}(t) + \sum_{\rho}\boldsymbol{\beta}_{RF\rho\mu}(t)\cdot \boldsymbol{z}_{RF\rho}^{(i)}\right)\prod_{\nu=1}^{\mu-1}\left(1-G_{\nu}^{(i)}\right)G_{\mu}^{(i)} + \beta_{PG}(t)\,x_P^{(i)}\right), (1)$$

$\lambda_0(t)$ is the baseline incidence. $G_{\mu}^{(i)}$ are indicator variables for the presence/absence of a PV in a major gene (MG), with $\mu = 1, \ldots, 5$ representing *BRCA1*, *BRCA2*, *RAD51D*, *RAD51C* and *BRIP1* respectively taking values 1 if a PV is present and 0 otherwise, and $\mu = 6$ corresponds to non-harbourer of PVs with $G_6^{(i)} = 1$ for non-harbourer, and 0 otherwise. $\beta_{MG\mu}(t)$ represent the age-specific log-relative risk (log-RR) associated with PV in the MGs, relative to the baseline incidence, with $\beta_{MG6}(t) = 0$. $x_P^{(i)}$ is the polygenotype for individual i, assumed to follow a normal distribution in the general population with mean 0 and variance 1, and $\beta_{PG}(t)$ is the age-specific log-RR associated with the polygene, relative to the baseline incidence. $\beta_{RF\rho\mu}(t)$ is the vector of the log-RRs associated with RF ρ at age t, which may depend on the major genotype, μ, and $\boldsymbol{z}_{RF\rho}^{(i)}$ is the corresponding indicator vector showing the category of RF ρ for the individual. The baseline incidence was determined by constraining the overall incidences to agree with the population EOC incidence. To allow

appropriately for missing RF information, only those RFs measured on a given individual are considered.

Major gene effects

To include the effects of *RAD51D, RAD51C* and *BRIP1*, we used the approach described in[31] where PVs in the three genes were assumed to be risk alleles of a single MG locus. To define the penetrance, we assumed the following order of dominance when an individual harboured more than one PV (that is, the risk was determined by the higher risk PV and the lower risk PV ignored): *BRCA1, BRCA2, RAD51D, RAD51C, BRIP1*[31]. The population allele frequencies for *RAD51D*, *RAD51C* and *BRIP1* and EOC relative risks were obtained from published data (Table s4)[15,29].

Epidemiological Risk Factors (RFs)

The RFs incorporated into the model include parity, use of oral contraception and MHT, endometriosis, tubal ligation, BMI and height (collectively referred to as RFs). We assumed that the RFs were categorical and that an individual's category was fixed for her remaining lifetime, although we allowed for the relative risks to vary with age. The RR estimates used in equation (1) and population distributions for each RF were obtained from large-scale external studies and from national surveillance data sources, using a synthetic approach as previously described[30]. Where possible, we used RR estimates that were adjusted for the other RFs included in the model, and distributions from the UK. Details of the population distributions and relative risks used in the model are given in the Tables s3. As in BOADICEA[30], in order to decrease the runtime, we combined the RFs that have age-independent relative risks into a single factor (specifically parity, tubal ligation, endometriosis, BMI and height).

Incorporating Polygenic Risk Scores

We included an EOC susceptibility PRS, assumed to form part of the polygene, using the methods previously developed[11,30]. The polygenic component decomposes into a measured component due to the PRS (x_{PRS}) and an unmeasured component representing other familial effects (x_R):

$$x_P = x_{PRS} + x_R. \tag{2}$$

x_{PRS} summarises the effects of multiple common variants and is assumed normally distributed with mean 0 and variance α^2 in the general population, with $0 \leq \alpha \leq 1$. The parameter α^2 is the proportion of the overall polygenic variance in model (after excluding the effects of all MGs) explained by the PRS. x_R is normally distributed with mean 0 and variance $1 - \alpha^2$. The approach used to calculate α^2 is described in the Supplementary Material. This implementation allows the effect size of the PRS to be dynamically varied, allowing for an arbitrary PRS. Here, we considered the latest validated EOC PRS developed by the Ovarian Cancer Association Consortium[32], which is composed of 36 variants (Table s2) and has a log-variance of 0.099, accounting for 5.0% of the overall polygenic variance in the model.

Other model components

The previous version[11] modelled the incidence of EOC and first female breast cancer (BC). To align with BOADICEA[30], the model was extended to take account of female contralateral BC, male BC, prostate and pancreatic cancer (Supplementary Methods).

Model Validation

Study subjects

The model validation was carried out in a nested case-control sample of females of self-reported White-European ancestry participating in UKCTOCS. Based on the data available, we were able to validate the model on the basis of family history, polygenic risk score and epidemiological risk factors. Details of the UKCTOCS study design, blood sampling process, DNA extraction and processing, variant selection, genotyping and data processing are described in the Supplementary Methods and published elsewhere[33]. In summary, the following self-reported information was collected at recruitment and used in the model validation: parity, use of oral contraception and MHT, tubal ligation, BMI and height (Table s5). Data on 15 common EOC susceptibility variants were also available (Table s6). Not all 36 currently known variants were available for the study sample, but as indicated earlier, the model can accommodate an arbitrary PRS. For this purpose, we used a 15-variant PRS[33], for which $\alpha^2 = 0.037$ (compared with $\alpha^2 = 0.05$ for the 36 variant PRS). Study participants were not screened for PVs in *BRCA1, BRCA2, RAD51C, RAD51D* or *BRIP1.*

Pedigree construction

The UKCTOCS recruitment questionnaire collected only summary data on FH of breast and ovarian cancer. Since the risk algorithm uses explicit FH information, the summary FH data were used to reconstruct the pedigrees, which included information on cancer occurrence in the first- and second-degree relatives (details in Supplementary Methods).

Statistical analysis

All UKCTOCS participants were followed using electronic health record linkage to national cancer and death registries. For this study, they were censored at either: their age at EOC,

their age at other (non-EOC) first cancer diagnosis, their age at death or age 79. To assess the model performance a weighted approach was used whereby each participant was assigned a sampling weight based on the inverse of the probability of being included in the nested case-control study, given their disease status. Since all incident cancer cases were included, cases were assigned a weight of 1. The cases were matched to two random controls (women with no EOC cancer) recruited at the same regional centre, age at randomisation and year at recruitment.

We assessed the model calibration and discrimination of the predicted five-year risks. Females older than 74 years at entry were excluded. Cases who developed EOC beyond five years were treated as unaffected. For controls with a less than five-year follow-up, we predicted the EOC risks to the age at censoring. For all other controls and cases, we predicted the EOC risks over five years.

To assess model calibration, we partitioned the weighted sample into quintiles of predicted risk. Within each quintile, we compared the weighted mean of predicted risk to the weighted observed incidence using the Hosmer-Lemeshow (HL) chi-squared test[34]. To assess relative risk calibration, the relative predicted and observed risks were calculated relative to the corresponding means of risks over all quintiles. We also compared the expected (E) with the observed (O) EOC risk within the prediction interval by calculating the ratio of expected to observed EOC cases (E/O). The 95% confidence interval (CI) for the ratio was calculated assuming a Poisson distribution[35].

We assessed the model discrimination between females who developed and did not develop EOC within five years using the area under the ROC curve (AUC). Details are given in the supplementary material.

Results

Model description

RAD51D, RAD51C and *BRIP1,* based on the assumed allele frequencies and relative risks, account for 2.5% of the overall polygenic variance in the model. Figure 1 shows the predicted EOC risk for harbourers of PVs in *BRCA1*, *BRCA2*, *RAD51D*, *RAD51C* and *BRIP1* for various FH scenarios. With unknown FH, the risks are 13%, 11% and 6% respectively. For example, for a *BRIP1* PV harbourer, the risk varies from 6% for a female without EOC FH to 18% for a female with two affected first-degree relatives. The model can also be used to predict risks in families in which PVs are identified, but where other family members test negative (Figure s1). For females with a family history of EOC, the reduction in EOC risk after negative predictive testing is greatest when a PV was identified in *BRCA1* in the family, with the risks being close to (though still somewhat greater than) population risk. This effect was most noticeable for females with a strong FH. Although a risk reduction is also seen for females whose mother harboured a PV in *BRCA2*, *RAD51D*, *RAD51C* or *BRIP1*, the reduction is less marked. As expected, the predicted risks are still elevated compared to population risk.

Figures 2 and s2 show distributions of lifetime risk and risk by age 50 respectively for females untested for PVs, predicted based on RFs and PRS, for two FH scenarios: (1) unknown FH (i.e. equivalent to a random female from the general population); and (2) having a mother diagnosed with EOC at age 50. Table 1 shows the corresponding proportion of females falling into different risk categories. The variation in risk is greatest when including both the RFs and PRS in the model. When considered separately, the distribution is widest for the RFs. Using the combined RF and PRS distribution, the predicted lifetime risks vary from 0.5% for the 1st

percentile to 4.6% for the 99th for a female with unknown FH and from 1.9% to 10.3% for a female with an affected mother.

Figure 3 shows the predicted lifetime EOC risk for harbourers of PVs in *BRCA1, BRCA2, RAD51D*, *RAD51C* and *BRIP1* based on RFs and PRS, for two FH scenarios. Taking a *RAD51D* PV harbourer for example, based on PV testing and FH alone, the predicted risks are 13% when FH is unknown and 23% when the female harbourer has a mother diagnosed with EOC at age 50. When RFs and the PRS are considered, the risks vary from 4% for those at the 1st percentile to 28% for those at the 99th percentile with unknown FH, and from 9% to 43% with an affected mother. Table 1 shows the proportion of females with PVs falling into different risk categories. Based on the combined risk distribution, 33% of *RAD51D* PV harbourers in the population are expected to have a lifetime EOC risk of less than 10%. Similarly, the distributions of risk for *BRIP*1 PV harbourers are shown in Figure 3(i) and 3(j) and in Table 1. Based on the combined RFs and PRS distributions, 46% of *BRIP1* PV harbourers in the population are expected to have lifetime risks of less than 5%; 47% to have risks between 5 and 10%, and 7% to have risks of 10% or greater. A *BRIP1* PV harbourer with an affected mother, on the basis of FH alone, has a lifetime risk of 11%. However, when the RFs and PRS are considered, 50% of those would be reclassified as having lifetime risks of less than 10%.

Figures s4 and s5 show the probability trees describing the reclassification of females as more information (RFs, PRS, and testing for PVs in the MGs) is added to the model for a female with unknown FH and a female with a mother affected at age 50 respectively based on the predicted lifetime risks. Figures s4(a) and s5(a) show the reclassification resulting from adding RFs, MG and PRS sequentially, while Figures s4(b) and s5(b) assume the order RFs, PRS and then MG. Assuming the three risk categories for lifetime risks are less than 5%, 5% or greater

but less than 10% and 10% or greater, there is significant risk reclassification as more information is added.

Model validation

After the censoring process, 1961 participants with 374 incident cases and 1587 controls met the 5-year risk prediction eligibility criteria. Table s5 summarises the characteristics of these cases and controls at baseline.

The model considering FH, the 15-variant PRS and a subset of the RFs (but not including testing the PV in the MGs) demonstrated good calibration in both absolute and relative predicted risk (Figure 4). Over the five-year period, the model predicted 391 EOCs, close to the 374 observed (E/O=1.05, 95%CI=0.94-1.16). The model was well calibrated across the quintiles of predicted risk (HL p=0.08), although there was a suggestion for an underprediction of risk in the lowest quintile (absolute risk E/O=0.66, 95%CI=0.52-0.91; RR E/O=0.63, 95%CI:0.42-0.95). The AUC for assessing discrimination of the full model was 0.61 (95%CI:0.58-0.64).

When looking at individual risk factors, FH predicted the widest 5-year risk variability (sd=0.0013; range 0.04% to 4.0%), followed by RFs (sd=0.0010; range:0.02%-0.7%) and PRS (sd=0.0009; range:0.05%-1.0%, Figure s6). As expected, their sequential inclusion increased the variability in the predicted risks (sd=0.0018; Figure s6).

Discussion

The EOC risk prediction model presented here combines the effects of FH of cancer, the explicit effects of rare moderate- to high-risk PVs in five established EOC susceptibility genes, a 36-variant PRS and other clinical and lifestyle factors (Table s1). The model provides a

consistent approach for estimating EOC risk on the basis of all known EOC risk factors and allows for prevention approaches to be targeted at those at highest risk.

The results demonstrate that in the general population, the existing PRS and RF alone identify 0.6% of females (with unknown family history) who have a lifetime risk of >5% (Table 1). On the other hand, in females with a positive family history, 37.1% of females would have a predicted risk between 5-10% and 1.2% would have an EOC risk of 10% or greater (Table 1). The results show that the RFs provide a somewhat greater level of risk stratification than the 36-variant PRS. However, risk discrimination is greater when the RFs and PRS are considered jointly. These results were in line with the observed risk distributions in the validation dataset, but direct comparisons are not possible due to the different variants included in the PRS used and limited RFs in the validation study. The results also show that significant levels of risk re-categorisation can occur for harbourers of PVs in the moderate or high-risk EOC susceptibility genes.

The comprehensive risk model is based on a synthetic approach previously used for breast cancer[30] and makes several assumptions. In particular, we assumed that the risks associated with known RFs and the PRS combine multiplicatively. We have not assessed this assumption in the present study, however, published studies found no evidence of deviations from the multiplicative model for the combined effect of the RFs and the PRS[28] suggesting that this assumption is reasonable. The underlying mathematical model is flexible enough to allow for deviations from this assumption should additional evidence become available.

Similarly, the model assumes that the relative effect-sizes of RFs and the PRS are similar in females harbouring PVs in *BRCA1, BRCA2, RAD51C, RAD51D* and *BRIP1* and those without PVs in these genes. Evidence from studies of *BRCA1* and *BRCA2* PV harbourers suggests that this

assumption is plausible: PRSs for EOC have been shown to be associated with similar relative risks of EOC in the general population and in *BRCA1* and *BRCA2* PV harbourers[36,37]. The current evidence also suggests that known epidemiological RFs have similar effect sizes for EOC in *BRCA1* and *BRCA2* PV harbourers as in non-harbourers[38,39]. No studies have so far assessed the joint effects of *RAD51C, RAD51D* and *BRIP1* PVs with the PRS, but the observation that FH modifies EOC risk for *RAD51C/D* PV harbourers[29] suggests that similar arguments are likely to apply. Large prospective studies are required to address these questions in more detail. We were not able to validate these assumptions in UKCTOCS because panel testing data were not available.

Other RFs for EOC that have been reported in the literature include breast feeding[26] and age at menarche and menopause[23]. However, the evidence for these risk factors is still limited. Our model is flexible enough to allow for additional RFs to be incorporated in the future.

We validated the five-year predicted risks on the basis of family history, risk factors and PRS in an independent dataset from a prospective trial[2]. A key strength was that EOC was a primary outcome in UKCTOCS. All women from the 202,638 randomised who developed EOC during a median follow-up of 11 years were included as cases. All cases were reviewed and confirmed by an independent outcome review committee[2]. The results indicated that absolute and relative risks were well-calibrated overall and in the top quintiles of predicted risk. However, there was some underprediction of EOC in the bottom quintile. This could be due to differences in the RF distributions in those who volunteer to participate in research (self-selected more healthy individuals[40]) compared to the general population, or due to random variations in the effects of the RFs in UKCTOCS compared to other studies. Alternatively, the multiplicative assumption may break down in the lowest risk category.

Further large prospective cohorts will be required to determine whether the underprediction in the lowest risk category reflects a systematic miscalibration of the model or is due to chance.

The current validation study has some limitations. The underlying model accounts for FH information on both affected and unaffected family members, but the UKCTOCS recruitment questionnaire did not include information on unaffected family members. Family sizes and ages for unobserved family members were imputed using demographic data. In addition, since data on whether the affected family members were from the paternal or maternal side, we assumed all the affected family members were from the same (maternal) side. These may result in inaccuracies in risk predictions. A further limitation is that UKCTOCS was undertaken to assess screening of low-risk women and therefore is not necessarily representative of a true population cohort of females, as females with a family history of two or more relatives with ovarian cancer or who were known harbourers of *BRCA1/2* PVs were not eligible to participate in the randomised controlled trial. Data were not available on the rare moderate and high-risk PVs, and we were only available to assess a PRS with 15-variants, rather than the more informative 36-variant implemented in the comprehensive model. Therefore, it has not been possible to validate the full model presented here. Future analyses in other cohorts will be required to further validate the full model.

In summary, we have developed the first comprehensive EOC risk prediction model that considers the currently known genetic and epidemiologic RFs and explicit FH of cancer. It allows users to obtain consistent, individualised EOC risks to better guide risk management. It can also be used to identify target populations for studies to assess novel prevention strategies (such as salpingectomy) or early detection approaches, by identifying those at

higher risk of developing the disease for enrolment into such studies. Future, independent studies should aim to validate the full model, including the full PRS and rare PVs in diverse settings. The model is available via the CanRisk Tool (www.canrisk.org), a user-friendly web tool that allows users to obtain the future risks of developing EOC easily.

Acknowledgements

This work has been supported by grants from Cancer Research UK (C12292/A20861).

The analysis is part of PROMISE, which was funded through Cancer Research UK PRC Programme Grant A12677 and by The Eve Appeal. University College London investigators received support from the National Institute for Health Research University College London Hospitals Biomedical Research Centre and from MRC core funding (MR_UU_12023). UKCTOCS was core funded by the Medical Research Council (G9901012 and G0801228), Cancer Research UK (C1479/A2884), and the Department of Health with additional support from the Eve Appeal.

FMW was supported by a National Institute for Health Research (NIHR) Clinician Scientist award (NIHR-CS-012-03).

MT was funded by the European Union Seventh Framework Program (2007–2013)/European Research Council (310018).

The authors are particularly grateful to the women throughout the UK who are participating in the trial and to the centre leads and the entire medical, nursing and administrative staff who work on the UKCTOCS.

Conflicts of Interest

The authors Douglas F. Easton, Antonis C. Antoniou, Alex P. Cunningham, Andrew Lee and Tim Carver, are listed as creators of the BOADICEA model, which has been licensed to Cambridge Enterprise for commercialisation.

Usha Menon has shares in Abcodia awarded to her by UCL.

References

1. Cancer Research UK. Ovarian Cancer Statistics. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer. Accessed 21 October, 2020.
2. Jacobs IJ, Menon U, Ryan A, et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet.* 2016;387(10022):945-956.
3. Rosenthal AN, Fraser LSM, Philpott S, et al. Evidence of Stage Shift in Women Diagnosed With Ovarian Cancer During Phase II of the United Kingdom Familial Ovarian Cancer Screening Study. *J Clin Oncol.* 2017;35(13):1411-1420.
4. Ludwig KK, Neuner J, Butler A, Geurts JL, Kong AL. Risk reduction and survival benefit of prophylactic surgery in BRCA mutation carriers, a systematic review. *Am J Surg.* 2016;212(4):660-669.
5. Manchanda R, Legood R, Antoniou AC, Gordeev VS, Menon U. Specifying the ovarian cancer risk threshold of 'premenopausal risk-reducing salpingo-oophorectomy' for ovarian cancer prevention: a cost-effectiveness analysis. *J Med Genet.* 2016;53(9):591-599.
6. Manchanda R, Legood R, Pearce L, Menon U. Defining the risk threshold for risk reducing salpingo-oophorectomy for ovarian cancer prevention in low risk postmenopausal women. *Gynecol Oncol.* 2015;139(3):487-494.
7. Parker WH, Feskanich D, Broder MS, et al. Long-term mortality associated with oophorectomy compared with ovarian conservation in the nurses' health study. *Obstet Gynecol.* 2013;121(4):709-716.

8. Rocca WA, Gazzuola-Rocca L, Smith CY, et al. Accelerated Accumulation of Multimorbidity After Bilateral Oophorectomy: A Population-Based Cohort Study. *Mayo Clin Proc.* 2016;91(11):1577-1589.

9. Parker WH, Jacoby V, Shoupe D, Rocca W. Effect of bilateral oophorectomy on women's long-term health. *Womens Health (Lond).* 2009;5(5):565-576.

10. Gaba F, Manchanda R. Systematic review of acceptability, cardiovascular, neurological, bone health and HRT outcomes following risk reducing surgery in BRCA carriers. *Best Pract Res Clin Obstet Gynaecol.* 2020;65:46-65.

11. Jervis S, Song H, Lee A, et al. A risk prediction algorithm for ovarian cancer incorporating BRCA1, BRCA2, common alleles and other familial effects. *J Med Genet.* 2015;52(7):465-475.

12. Loveday C, Turnbull C, Ramsay E, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat Genet.* 2011;43(9):879-882.

13. Loveday C, Turnbull C, Ruark E, et al. Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat Genet.* 2012;44(5):475-476; author reply 476.

14. Pal T, Akbari MR, Sun P, et al. Frequency of mutations in mismatch repair genes in a population-based study of women with ovarian cancer. *Br J Cancer.* 2012;107(10):1783-1790.

15. Ramus SJ, Song H, Dicks E, et al. Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. *J Natl Cancer Inst.* 2015;107(11).

16. Yang X, Leslie G, Doroszuk A, et al. Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *J Clin Oncol.* 2020;38(7):674-685.

17. Kuchenbaecker KB, Ramus SJ, Tyrer J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nature Genetics.* 2015;47(2):164-171.

18. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet.* 2017.

19. Collaborative Group on Epidemiological Studies of Ovarian Cancer. Ovarian cancer and body size: individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS Med.* 2012;9(4):e1001200.

20. Cancer CGoESoO. Menopausal hormone use and ovarian cancer risk: individual participant meta-analysis of 52 epidemiological studies. *The Lancet.* 2015;385(9980):1835-1842.

21. Li K, Husing A, Fortner RT, et al. An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study. *Br J Cancer.* 2015;112(7):1257-1265.

22. Pearce CL, Stram DADO, Ness RB, et al. Population distribution of lifetime risk of ovarian cancer in the United States. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2015;24(4):671-676.

23. Rosner BA, Colditz GA, Webb PM, Hankinson SE. Mathematical models of ovarian cancer incidence. *Epidemiology (Cambridge, Mass).* 2005;16(4):508-515.

24. Collaborative Group on Epidemiological Studies of Ovarian Cancer, Beral V, Gaitskell K, et al. Ovarian cancer and smoking: individual participant meta-analysis including 28,114 women with ovarian cancer from 51 epidemiological studies. *The Lancet Oncology.* 2012;13(9):946-956.

25. Gaitskell K, Coffey K, Green J, et al. Tubal ligation and incidence of 26 site-specific cancers in the Million Women Study. *British journal of cancer.* 2016;114(9):1033-1037.

26. Jordan SJ, Cushing-Haugen KL, Wicklund KG, Doherty JA, Rossing MA. Breast-feeding and risk of epithelial ovarian cancer. *Cancer causes & control : CCC.* 2012;23(6):919-927.

27. Pfeiffer RM, Park Y, Kreimer AR, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. *PLoS Med.* 2013;10(7):e1001492.

28. Clyde MA, Weber RP, Iversen ES, et al. Original Contribution Risk Prediction for Epithelial Ovarian Cancer in 11 United States–Based Case-Control Studies: Incorporation of Epidemiologic Risk Factors and 17 Confirmed Genetic Loci. *American Journal of Epidemiology.* 2016.

29. Yang X, Song H, Leslie G, et al. Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. *J Natl Cancer Inst.* 2020.

30. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med.* 2019;21(8):1708-1718.

31. Lee A, Cunningham AP, Tischkowitz M, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. *Genet Med.* 2016;18(12):1190-1198.

32. Dareng EO, Tyrer J, Barnes DR, et al. Polygenic Risk Modelling for Prediction of Epithelial Ovarian Cancer Risk. *medRxiv.* 2020:2020.2011.2030.20219220.

33. Yang X, Leslie G, Gentry-Maharaj A, et al. Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J Med Genet.* 2018;55(8):546-554.

34. Whittemore AS, Halpern J. Two-stage sampling designs for external validation of personal risk models. *Statistical Methods in Medical Research.* 2016;25(4):1313-1329.

35. Terry MB, Liao Y, Whittemore AS, et al. 10-year performance of four models of breast cancer risk: a validation study. *The Lancet Oncology.* 2019;20(4):504-517.

36. Barnes DR, Rookus MA, McGuffog L, et al. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med* 2020;in Press.

37. Kuchenbaecker KB, McGuffog L, Barrowdale D, et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst.* 2017;109(7).

38. Antoniou AC, Rookus M, Andrieu N, et al. Reproductive and hormonal factors, and ovarian cancer risk for BRCA1 and BRCA2 mutation carriers: results from the International BRCA1/2 Carrier Cohort Study. *Cancer Epidemiol Biomarkers Prev.* 2009;18(2):601-610.

39. Milne RL, Antoniou AC. Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. *Endocr Relat Cancer.* 2016;23(10):T69-84.

40. Burnell M, Gentry-Maharaj A, Ryan A, et al. Impact on mortality and cancer incidence rates of using random invitation from population registers for recruitment to trials. *Trials.* 2011;12:61.

Figure Legends

Figure 1. Predicted lifetime (age 20 to 80 years) EOC risk by PV and family history. Each Figure shows the risks assuming the female is untested, has no PVs, or harbours a PV in *BRCA1*, *BRCA2*, *RAD51D*, *RAD51C* or *BRIP1*. Figure (a) assumes unknown family history. Figures (b)-(e) assume an increasing number of affected first-degree relatives, as indicated by the pedigree diagram inserts. Predictions are based on UK EOC population incidence.

Figure 2. Predicted lifetime (age 20 to 80 years) EOC risk for a female untested for PVs based on the different predictors of risk (risk factors (RFs) and PRS). Figures (a) and (c) show the risk for a female with an unknown family history (equivalent to the distribution of risk in the population), while figures (b) and (d) show the risk for a female with a mother affected at age 50. Figures (a) and (b) show the probability density function against absolute risk, while figures (c) and (d) show absolute risk against cumulative distribution. The vertical line in figure (a) and the horizontal line in Figure (c) (labelled "no RFs or PRS") is equivalent to the population risk of EOC. The "Population" risk is shown separately in figures (b) and (d). Predictions are based on UK EOC population incidences.

Figure 3. Predicted lifetime EOC risk for a female who has a PV in one of the high- or intermediate-risk genes included in the model, based on the different predictors of risk (risk factors (RFs) and PRS), for two family histories. Figures (a) and (b) show the lifetime risk for a harbourer of a PV in *BRCA1*, Figures (c) and (d) show the lifetime risk for a harbourer of a PV in *BRCA2*, Figures (e) and (f) show the lifetime risk for a harbourer of a PV in *RAD51D*, Figures (g) and (h) show the lifetime risk for a harbourer of a PV in *RAD51C*, while Figure (i) and (j) show the lifetime risk for a harbourer of a PV in *BRIP1*. Figures (a), (c), (e), (g) and (i) show risks for an unknown family history, while Figures (b), (d), (f), (h) and (j) show risks for a female

whose mother is diagnosed with EOC at age 50. Predictions based on UK ovarian cancer incidences.

Figure 4. Calibration of the absolute and relative predicted five-year EOC risks, showing the observed and expected risks by quintile. The bars show the 95% confidence intervals for the observed risks. Relative risks were calculated relative to the overall mean of observed and predicted risks.

Tables

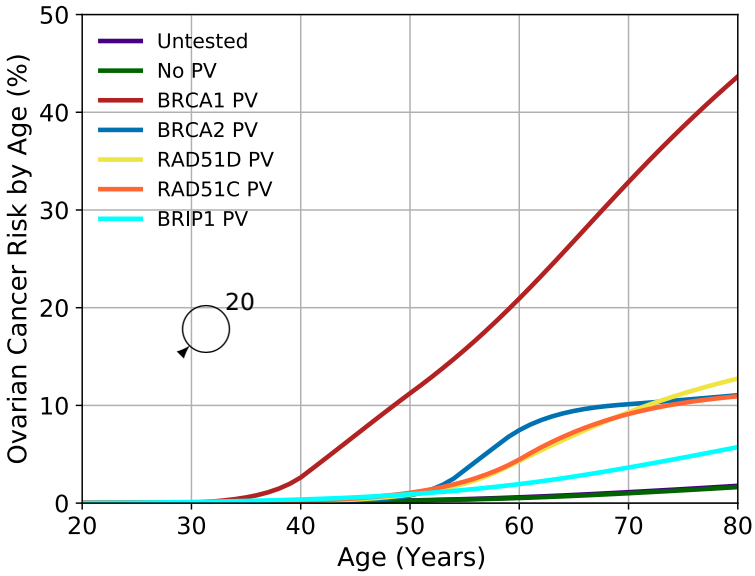
PV harbourer status	Family history	Risk Categories	Lifetime Risk*			Risk to age 50*			Ref Fig.
			RF	PRS	RF and PRS	RF	PRS	RF and PRS	
Untested	NA	population	99.9	100.0	99.4	100.0	100.0	100.0	2(a)(c) and s2(a)(c)
		moderate	0.1	0.0	0.6	0.0	0.0	0.0	
		high	0.0	0.0	0.0	0.0	0.0	0.0	
	M 50	population	59.1	61.8	61.7	100.0	100.0	100.0	2(b)(d) and s2(b)(d)
		moderate	40.3	38.2	37.1	0.0	0.0	0.0	
		high	0.6	0.0	1.2	0.0	0.0	0.0	
***BRCA1* harbourer**	NA	population	0.0	0.0	0.0	0.0	0.0	0.3	3(a) and s3(a)
		moderate	0.0	0.0	0.0	1.5	0.1	4.1	
		high	100.0	100.0	100.0	98.5	99.9	95.6	
	M 50	population	0.0	0.0	0.0	0.0	0.0	0.0	3(b) and s3(b)
		moderate	0.0	0.0	0.0	0.0	0.0	0.2	
		high	100.0	100.0	100.0	100.0	100.0	99.8	
***BRCA2* harbourer**	NA	population	1.6	0.1	4.7	100.0	100.0	99.9	3(c) and s3(c)
		moderate	42.8	37.7	42.3	0.0	0.0	0.1	
		high	55.6	62.2	52.9	0.0	0.0	0.0	
	M 50	population	0.0	0.0	0.1	95.9	99.9	95.3	3(d) and s3(d)
		moderate	1.9	0.1	4.7	4.0	0.1	4.5	
		high	98.1	99.9	95.3	0.0	0.0	0.2	
***RAD51D* harbourer**	NA	population	0.6	0.0	2.1	100.0	100.0	99.8	
		moderate	24.5	18.3	30.8	0.0	0.0	0.2	

PV harbourer status	Family history	Risk Categories	Lifetime Risk*			Risk to age 50*			Ref Fig.
			RF	PRS	RF and PRS	RF	PRS	RF and PRS	
		high	75.0	81.7	67.1	0.0	0.0	0.0	3(e) and s3(e)
	M 50	population	0.0	0.0	0.0	93.5	99.6	92.6	3(f) and s3(f)
		moderate	0.6	0.0	2.0	6.3	0.4	7.0	
		high	99.4	100.0	98.0	0.2	0.0	0.4	
***RAD51C* harbourer**	NA	population	1.8	0.1	4.9	99.8	100.0	99.3	3(g) and s3(g)
		moderate	44.4	39.3	43.1	0.2	0.0	0.7	
		high	53.7	60.6	52.0	0.0	0.0	0.0	
	M 50	population	0.0	0.0	0.1	85.8	96.2	85.0	3(h) and s3(h)
		moderate	2.0	0.1	5.0	13.6	3.8	13.7	
		high	98.0	99.9	95.0	0.6	0.0	1.3	
***BRIP1* harbourer**	NA	population	43.6	34.0	45.7	100.0	100.0	99.7	3(i) and s3(i)
		moderate	52.3	65.0	47.4	0.0	0.0	0.3	
		high	4.1	1.0	6.9	0.0	0.0	0.0	
	M 50	population	2.1	0.1	4.8	93.0	99.3	91.4	3(j) and s3(j)
		moderate	43.5	38.9	44.5	6.8	0.7	8.1	
		high	54.5	61.1	50.7	0.3	0.0	0.5	

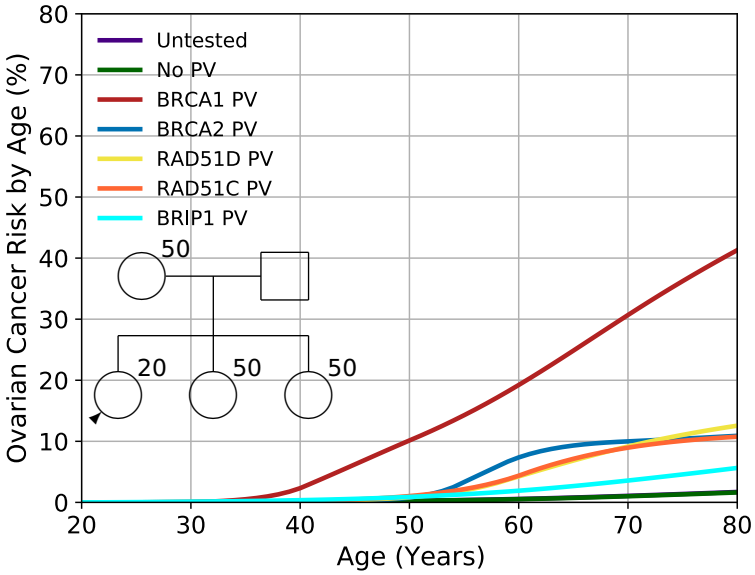
Table 1. Percentage of females falling in different risk categories by status of PV in one of the high- or intermediate-risk genes included in the model and family history of cancer. The "population" risk category is defined as lifetime risk of <5% and risk to age 50 of <3%. "Moderate" risk category is defined as a lifetime risk of 5% or greater but less than 10%[6] and a risk to age 50 of 3% or greater but

less than 5%. “High” risk category is defined as a lifetime risk of 10% or greater and a risk to age 50 of 5% or greater. RF: epidemiological risk factors; PRS: polygenic risk score; NA: unknown family history; M 50: mother diagnosed with EOC at age 50. The population lifetime risk is 1.8%, and the population risk to age 50 is 0.27%.

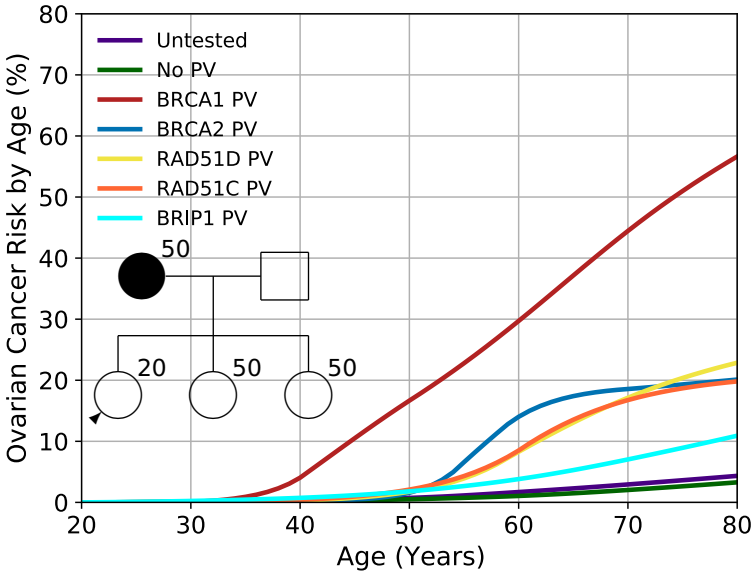
Figure 1



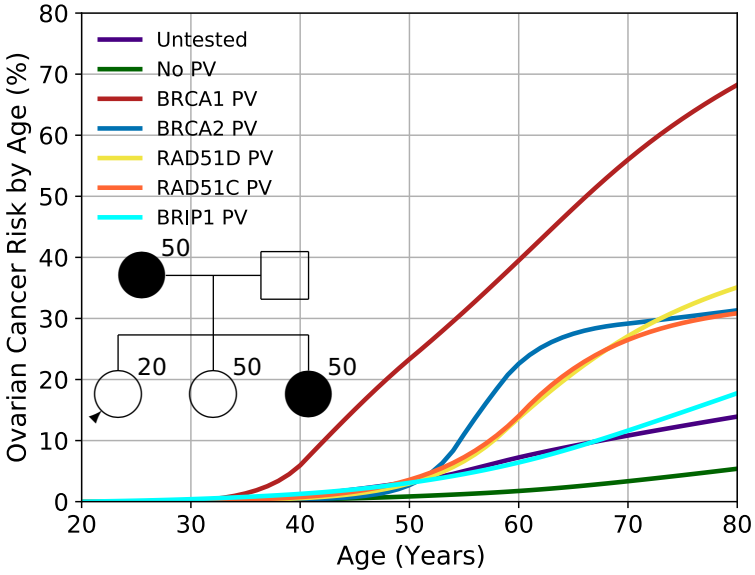
(a)



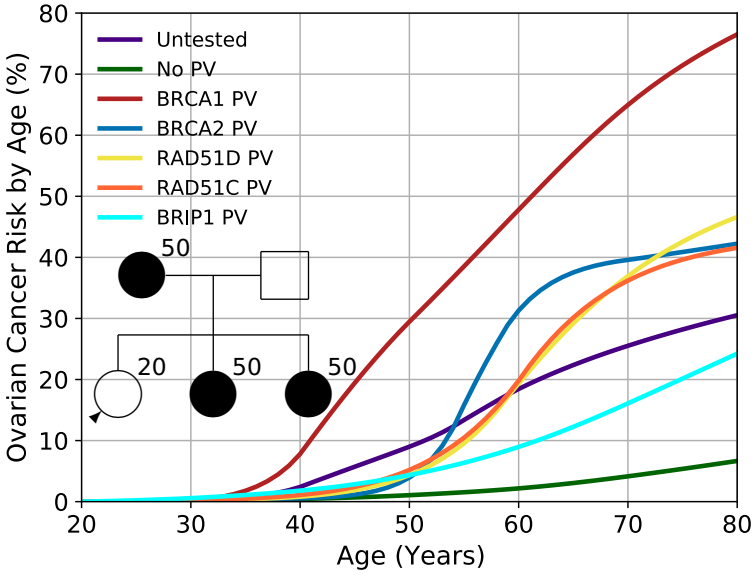
(b)



(c)



(d)



(e)

Figure 2

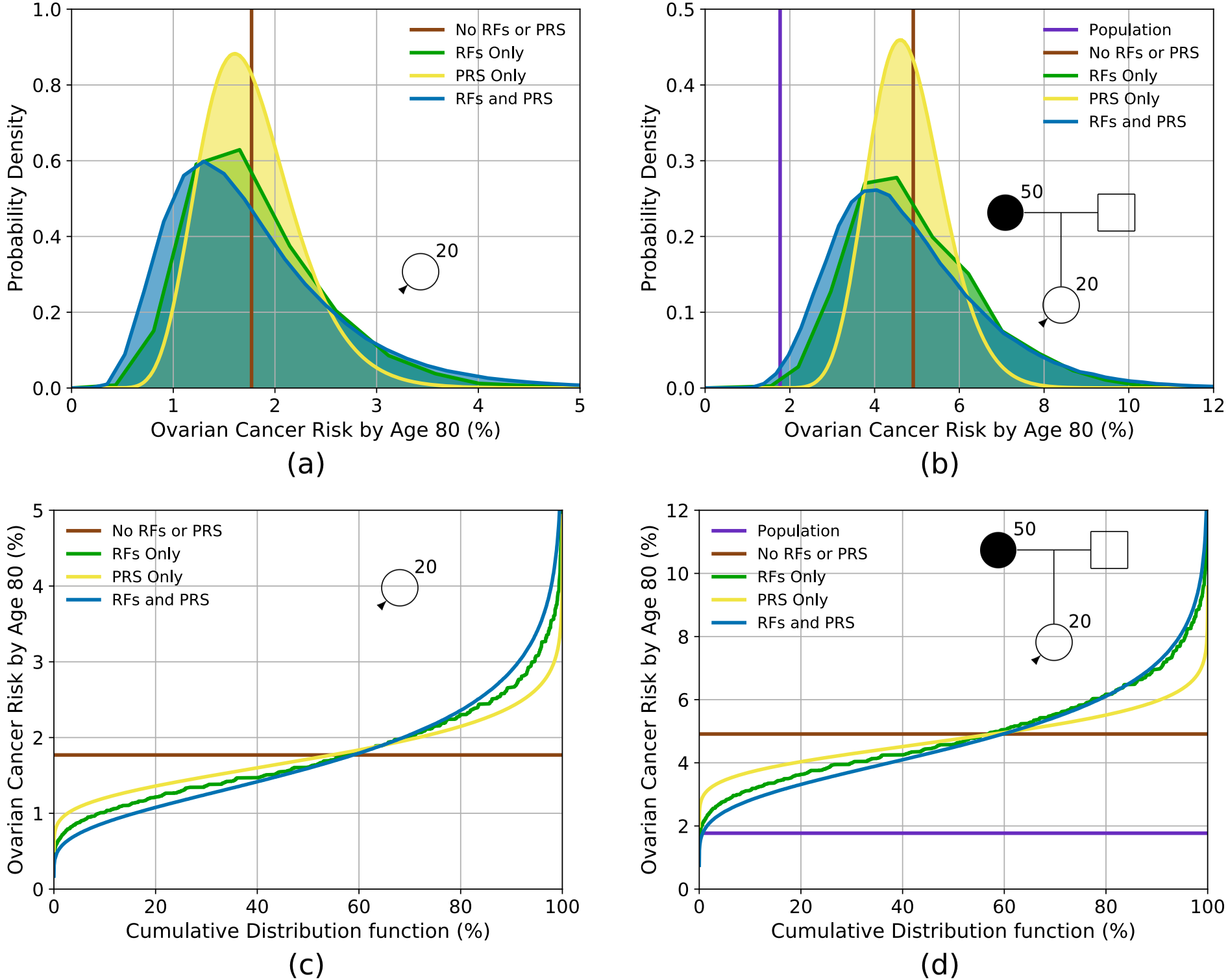

Probability Density
Ovarian Cancer Risk by Age 80 (%)
No RFs or PRS
RFs Only
PRS Only
RFs and PRS
20
(a)
Population
50
(b)
Ovarian Cancer Risk by Age 80 (%)
Cumulative Distribution function (%)
(c)
(d)

Figure 3

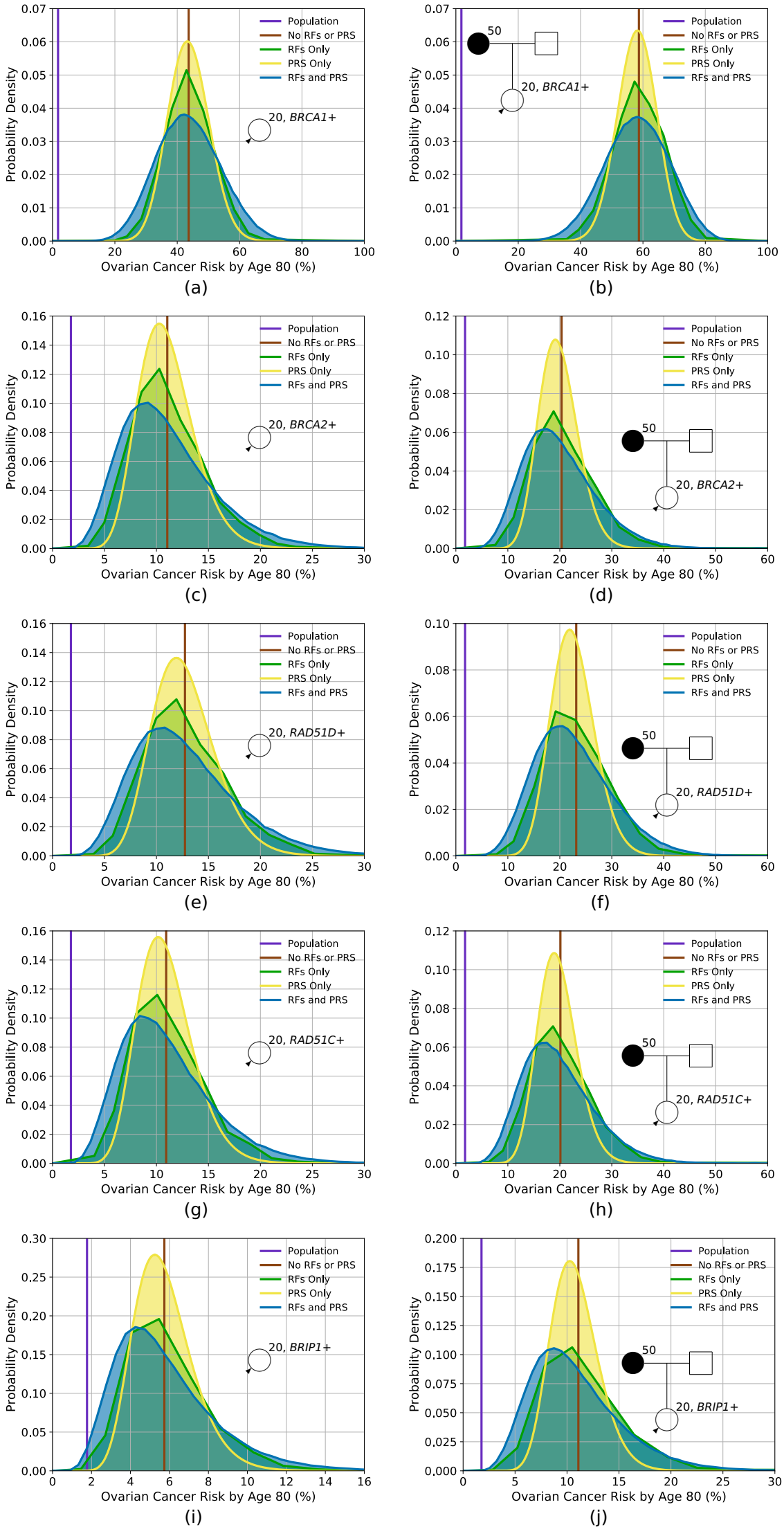
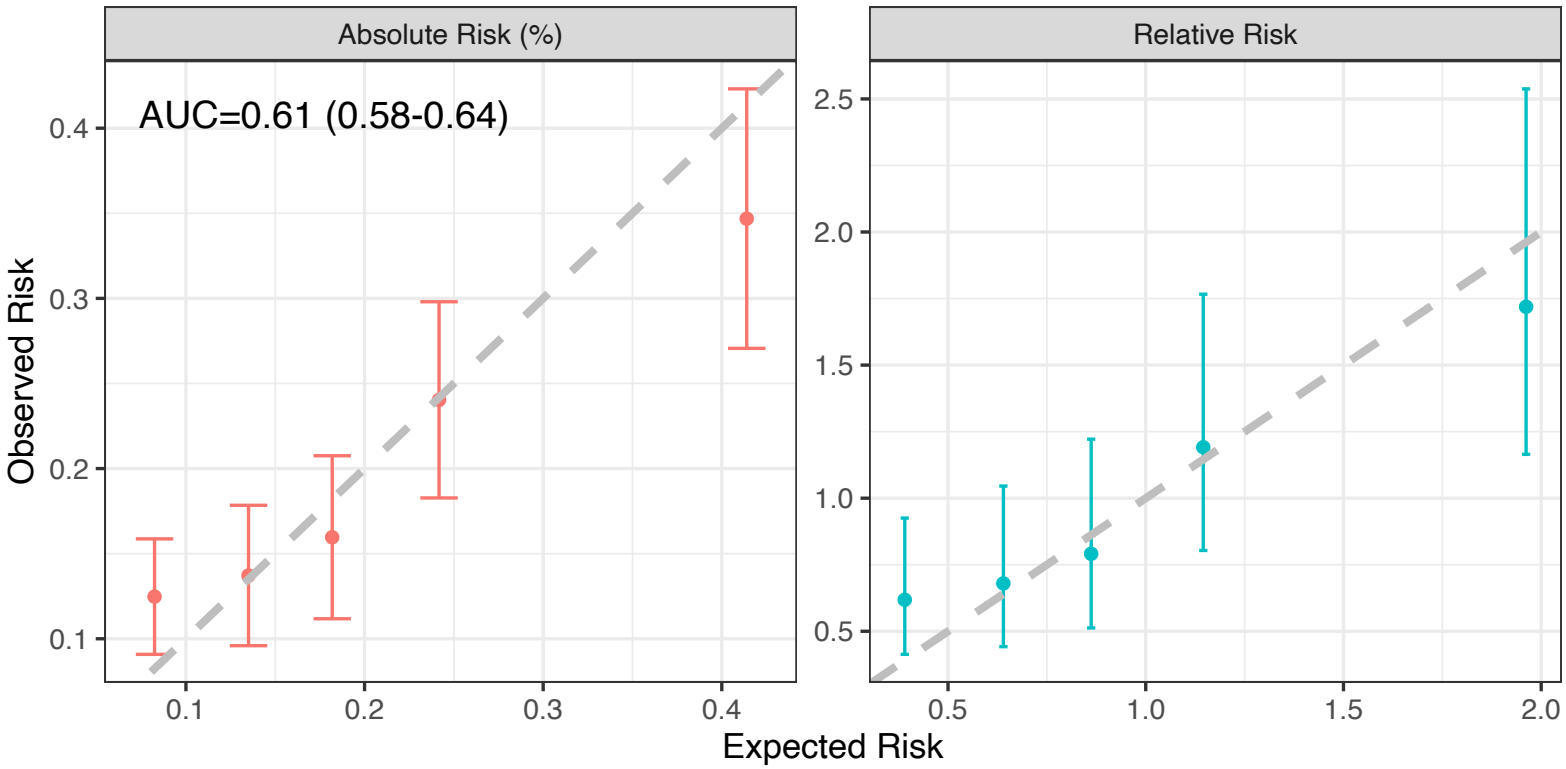

Population
No RFs or PRS
RFs Only
PRS Only
RFs and PRS
20, BRCA1+
50
20, BRCA2+
20, RAD51D+
20, RAD51C+
20, BRIP1+
Probability Density
Ovarian Cancer Risk by Age 80 (%)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)

Figure 4


Absolute Risk (%)
Relative Risk
AUC=0.61 (0.58-0.64)
Observed Risk
Expected Risk
0.1
0.2
0.3
0.4
0.5
1.0
1.5
2.0
2.5